For the Ukrainian language, there are some unfinished works (for example, "Free algorithm for stigma for the Ukrainian language" http://www.senyk.poltava.ua/projects/projects.html) or complex systems, but without the necessary functional (for example, "LanguageTool" https://www.languagetool.org/development/). Therefore, there is a need to implement their own libraries for processing and analyzing texts.

After analyzing the classic approaches to the text mining text analysis, we adapted Porter's algorithm and implemented a freely available PHP language module for tokenization and emulation of the Ukrainian language. Practical testing has shown that for tokenization, a simple breakdown into spaces with the subsequent normalization of tokens works well. For stemming, we used an approach that cuts the end and suffix, bringing words to its root as closely as possible. A better result can be achieved through lemmatization, but such a solution requires a separate study of the language and large resources.

The resulting initial processing of the text in this article allows one to work with tokens to obtain semantic links in texts written in Ukrainian. The stemmer designed as a module that can be quickly integrated into the project and available for use at the link https://packagist.org/packages/tochytskyi/ukrstemmer.

Keywords: text mining, tokenization, lematization, stemming, semantic text analysis.

Матеріал надійшов 12.09.2017

УДК 519.85,519.172

Стецюк П. І., Ляшко В. І., Бардадим Т. О.

ВЛАСТИВОСТІ КВАДРАТИЧНОЇ ЗАДАЧІ ПРО МАКСИМАЛЬНИЙ *k*-ПЛЕКС У НЕОРІЄНТОВАНОМУ ГРАФІ

У статті розглянуто властивості верхніх оцінок для квадратичної задачі про максимальний k-плекс у неорієнтованому графі. Проаналізовано зв'язок квадратичної задачі для 1-плекса з відомим формулюванням квадратичної задачі для знаходження максимальної кліки графа. Наведено лагранжеві двоїсті оцінки для найпростіших квадратичних задач та показано, що їх можна покращити при додаванні функціонально надлишкових обмежень.

Ключові слова: максимальний *k*-плекс, максимальна кліка, квадратична задача, лагранжева двоїста оцінка, функціонально надлишкове обмеження.

Вступ

Поняття *k*-плекса для неорієнтованого графа введено в [6] (*k* – деяке натуральне число). Якщо *k*=1, то *k*-плекс збігається з клікою (повним підграфом) графа. При $k > 1$ *k*-плекс є ослабленням поняття кліки графа і вимоги на включення вершини в *k*-плекс є слабшими, ніж вимоги на включення вершини в кліку. Ці поняття широко використовуються в соціології для виявлення та дослідження окремих підгруп населення, при кластеризації даних, для оптимізації інформаційних потоків у мережах тощо (див., напр., [2–5]). Слід зазначити, що зазвичай оптимізаційні задачі пошуку максимальних клік та *k*-плексів є NP-складними.

У статті буде проаналізовано зв'язок наведеного у [1] формулювання квадратичної задачі для 1-плекса з відомим формулюванням квадратичної задачі для знаходження максимальної кліки графа, буде розглянуто лагранжеві двоїсті оцінки для ряду квадратичних задач [7] та показано, що лагранжеві двоїсті оцінки можна покращити за допомогою додавання функціонально надлишкових обмежень, розглянутих у роботі [1]. Для обчислення лагранжевих двоїстих оцінок використано програму DSQTPr [8].

Означення та формулювання задачі

Нехай $G=(V,E)$ – неорієнтований граф із множиною вершин $V=\{1,\dots,n\}$ та множиною ребер *E*. Ребро графа *G*, що зв'язує вершини $i\in V$ та $j\in V$, умовимося позначати $(i,j)\in E$. Для графа *G* будемо використовувати також іншу форму його представлення: $G=(V,\Gamma)$, де $\Gamma=\{\Gamma(i),\ i=1,\dots,n\}$, а $\Gamma(i)$ – кінцеві вершини тих дуг, у яких початковою вершиною є вершина *i*. Кількість ребер графа *G* в обох представленнях зв'язана співвідношенням: $|E|=\frac{1}{2}\sum_{i\in V}|\Gamma(i)|$.

Комплементарний до *G* граф будемо позначати $\overline{G}=(V,\overline{E})$ і $\overline{G}=(V,\overline{\Gamma})$, де $(i,j)\in\overline{E}$ і $\overline{\Gamma}=\{\overline{\Gamma}(i),\ i=1,\dots,n\}$.

Підмножину вершин *S* із *V* називають *k*-плексом графа *G*, якщо ступінь кожної вершини в індукованому підграфі $G[S]$ (підграфі, породженому підмножиною *S*) є не меншим, ніж $|S|-k$. Тобто $S\subset V$ є *k*-плексом, якщо виконується така умова:

$$\deg_{G[S]}(i)=|\Gamma(i)\cap S|\ge|S|-k \qquad \forall i\in S.$$

k-плекс є максимальним по включенню (maximal), якщо він не міститься в жодному іншому *k*-плексі. Найбільший із максимальних по включенню *k*-плексів називають максимальним (maximum), його розмір називають *k*-плексним числом графа *G* та позначають $\rho_k(G)$ [2]. Очевидно, що 1-плекс є клікою графа *G*, тому що ступінь кожної вершини в індукованому підграфі $G[S]$ не менший, ніж $|S|-1$, а це означає, що кожна з вершин у підграфі $G[S]$ зв'язана з усіма іншими вершинами, тобто підграф $G[S]$ є повним підграфом (клікою) графа *G*. У цьому випадку $\rho_1(G)=\omega(G)$, де $\omega(G)$ – кліковe число графа *G*, що відповідає розміру максимальної кліки в графі *G*.

Для знаходження максимального *k*-плекса графа *G* у роботі [1] сформульовано оптимізаційну квадратичну задачу в такій формі:

$$\rho_k(G)=\max_x\sum_{i\in V}x_i \tag{1}$$

при обмеженнях:

$$\sum_{j\in\overline{\Gamma}(i)}x_ix_j\le(k-1)x_i,\quad \forall i\in V, \tag{2}$$

$$x_i^2-x_i=0,\ \ \forall i\in V. \tag{3}$$

Тут вершині $i\in V$ $(i=1,2,\dots)$ відповідає булева змінна $x_i\in\{0,1\}$ така, що $x_i=1$, якщо $i\in S$, та $x_i=0$, якщо $i\in V\setminus S$. Якщо у формулюванні задачі (1)–(3) квадратичні нерівності (2) замінити на квадратичні обмеження

$$\sum_{j\in\overline{\Gamma}(i)}x_ix_j=(k-1)x_i,\quad \forall i\in V,$$

то отримаємо формулювання квадратичної задачі для знаходження максимального «строгого» *k*-плекса графа *G*, тобто підмножини його вершин, для яких ступінь вершини дорівнює $|S|-k$.

Максимальний 1-плекс та максимальна кліка

Оскільки для графа *G* 1-плекс збігається з клікою, природно, що знаходження максимального 1-плекса графа *G* у задачі (1)–(3) відповідає знаходженню максимальної кліки графа *G*. Якщо *k*–1, то права частина нерівностей (2) дорівнює нулю, і задача (1)–(3) переходить у таку квадратичну оптимізаційну задачу:

$$\rho_1(G)=\omega(G)=\max\sum_{i\in V}x_i \tag{4}$$

при обмеженнях:

$$\sum_{j\in\overline{\Gamma}(i)}x_ix_j\le 0,\quad \forall i\in V, \tag{5}$$

$$x_i^2-x_i=0,\ \ \forall i\in V. \tag{6}$$

Задачу (4)–(6) можна розглядати як одне з можливих квадратичних формулювань задачі про знаходження максимальної кліки графа *G*. Задачі (4)–(6) відповідає рівно *n* квадратичних обмежень нерівностей (5).

Більш відомим квадратичним формулюванням задачі для максимальної кліки графа *G* є квадратична оптимізаційна задача [7]:

$$\omega(G)=\max\sum_{i\in V}x_i \tag{7}$$

при обмеженнях:

$$x_ix_j=0,\quad \forall(i,j)\in\overline{E}, \tag{8}$$

$$x_i^2-x_i=0,\ \ \forall i\in V. \tag{9}$$

Задачу (7)–(9) можна інтерпретувати як очевидний наслідок задачі (4)–(6), тому що всі змінні x_i, $i=1,\dots,n$ є невід'ємними. Дійсно, згідно з квадратичними обмеженнями (5), сума невід'ємних добутків пар змінних повинна бути не більше нуля, а це еквівалентно (за умови невід'ємності) системі квадратичних рівностей, де кожен окремий добуток двох змінних, котрі входять в обмеження (5), дорівнює нулю. Це й становить зміст обмежень (8).

Яка з квадратичних моделей є кращою для знаходження максимальної кліки графа *G*? Формально задача (4)–(6) має ту перевагу, що вона містить меншу кількість квадратичних обмежень, ніж задача (7)–(9). Однак вона програє в точності лагранжевої двоїстої оцінки.

Нехай $\psi^*_{\rho_1}(G)$ – оптимальна лагранжева оцінка для задачі (4)–(6), а $\psi^*_{\omega}(G)$ – оптимальна лагранжева оцінка для задачі (7)–(9). Кожна з них буде верхньою оцінкою для $\omega(G)$. Верхня оцінка $\psi^*_{\rho_1}(G)$ буде зазвичай гіршою за оцінку $\psi^*_{\omega}(G)$. Це можна пояснити тим, що в задачі (7)–(9) ми маємо більшу свободу у виборі множників Лагранжа, що відповідають обмеженням (8): кожному з ребер комплементарного графа $\overline{G}$ буде відповідати свій множник Лагранжа. Для обмежень (5) таких множників Лагранжа усього n, тобто свій множник Лагранжа буде відповідати кожній з вершин графа G. Крім того, свобода на вибір множників Лагранжа для обмежень (5) обмежена ще й тим, що для кожного з цих множників потрібно врахувати його невід'ємність. Урахування невід'ємності обмежує можливості вибору множників Лагранжа в задачі (4)–(6) у порівнянні з випадком, коли обмеження (5) замінюються на

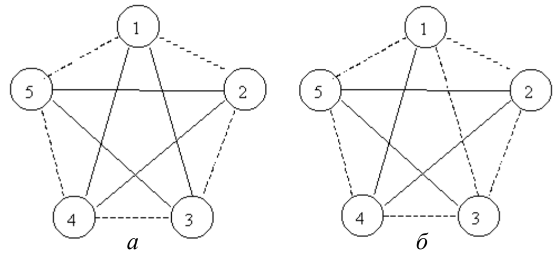
$$\sum_{j \in \overline{\Gamma}(i)} x_i x_j = 0, \quad \forall i \in V,$$

(вони випливають із обмежень (8), якщо їх згрупувати по вершинах графа G). Зауважимо, що квадратичну задачу з цими обмеженнями замість обмежень (5) можна розглядати як задачу знаходження максимального «строгого» 1-плекса графа G. Природно, що лагранжева двоїста оцінка для квадратичної задачі, що відповідає максимальному «строгому» 1-плексу графа G, буде в багатьох випадках точнішою за оцінку $\psi^*_{\rho_1}(G)$.

Проілюструємо поводження лагранжевих двоїстих оцінок $\psi^*_{\rho_1}(G)$ та $\psi^*_{\omega}(G)$ на прикладі двох графів G_1 і G_2 (рис. 1), що складаються усього з п'яти вершин. Тут ребра графів G_1 і G_2 позначено суцільною лінією, а ребра комплементарних графів $\overline{G_1}$ і $\overline{G_2}$ – пунктирною. Для графа G_1 (рис. 1, *а*) системи обмежень для задач (4)–(6) та (7)–(9) будуть такими:

$$\begin{array}{ll} x_1x_2 + x_1x_5 \le 0, & x_1x_2 = 0, \\ x_1x_2 + x_2x_3 \le 0, & x_2x_3 = 0, \\ x_2x_3 + x_3x_4 \le 0, & x_3x_4 = 0, \\ x_3x_4 + x_4x_5 \le 0, & x_4x_5 = 0, \\ x_4x_5 + x_1x_5 \le 0, & x_1x_5 = 0. \end{array}$$

Лагранжеві двоїсті оцінки $\psi^*_{\rho 1}(G)$ та $\psi^*_{\omega 1}(G)$ однакові й дорівнюють $\sqrt{5}$. Вони є неточними оцінками зверху, тому що $\omega(G_1) = 2$. Приклад для графа G_1 демонструє, що лагранжеві двоїсті оцінки $\psi^*_{\rho_1}(G)$ та $\psi^*_{\omega}(G)$ можуть співпадати.



а *б*

Рис. 1. Приклади графів G_1 та G_2

Це не виконується для графа G_2 на рис. 1, *б*, що відрізняється від G_1 тим, що з нього вилучено ребро (1,3). Для графа G_2 системи обмежень задач (4)–(6) та (7)–(9) будуть такими:

$$\begin{array}{ll} x_1x_2 + x_1x_3 + x_1x_5 \le 0, & x_1x_2 = 0, \\ x_1x_2 + x_2x_3 \le 0, & x_1x_3 = 0, \\ x_1x_3 + x_2x_3 + x_3x_4 \le 0, & x_2x_3 = 0, \\ x_3x_4 + x_4x_5 \le 0, & x_3x_4 = 0, \\ x_4x_5 + x_1x_5 \le 0. & x_4x_5 = 0, \\ & x_1x_5 = 0. \end{array}$$

Цього разу оцінки $\psi^*_{\rho_1}(G)$ та $\psi^*_{\omega}(G)$ різні. Зокрема, оцінка $\psi^*_{\omega}(G) = 2.0$ є точною верхньою оцінкою для $\omega(G_2)$. Оцінка $\psi^*_{\rho_1}(G) = 2.15222$ не є точною і потребує уточнення. Отже, приклад графа G_2 показує, що лагранжева двоїста оцінка для задачі (7)–(9) має перевагу над відповідною оцінкою задачі (4)–(6).

Зауважимо, що якщо для графа G_2 обмеження нерівності замінити на обмеження у формі рівностей, що відповідає знаходженню максимального «строгого» 1-плекса графа G_1, то лагранжева двоїста оцінка стане рівною двом і буде точною верхньою оцінкою для $\omega(G_1)$.

Якщо потрібно знайти більш точну лагранжеву двоїсту оцінку для максимальної кліки графа *G* за допомогою одної з двох розглянутих квадратичних задач, то краще користуватися квадратичною задачею у формі (7)–(9), ніж квадратичною задачею у формі (4)–(6), що для 1-плекса випливає з квадратичного формулювання (1)–(3). Аналогічна ситуація матиме місце і для квадратичної задачі [7], що пов'язана із знаходженням максимальної незалежної множини вершин графа.

Однак ситуація може змінитися, якщо мова йтиме не про знаходження оцінок $\psi^*_{\rho_1}(G)$ та $\psi^*_{\omega}(G)$, а про їх поліпшення за допомогою додавання функціонально надлишкових обмежень. Тоді квадратична задача (4)–(6) може мати переваги хоча б тому, що до неї можна додати більшу кількість функціонально надлишкових обмежень. Крім того, коли $k \geq 2$, то для знаходження максимального k-плекса навряд чи можна знайти альтернативу квадратичному формулюванню (1)–(3). Тому дослідження точності лагранжевих двоїстих оцінок для задачі (1)–(3), що підсилена за рахунок додавання функціонально надлишкових квадратичних обмежень, становить значний інтерес. У результаті може виявитися, що безпосереднє перенесення таких оцінок на окремий випадок $k = 1$ для ряду графів може привести до досить хороших за точністю верхніх оцінок для $\omega(G)$.

Для уточнення верхніх оцінок можна використовувати квадратичні нерівності з двох сімейств функціонально надлишкових квадратичних обмежень [7]. Вони побудовані за схемою, яку використовував Н. З. Шор для задачі про максимальну незалежну множину вершин графа [7, с. 250]. За основу береться формулювання задачі про максимальний k-плекс, яке має вигляд:

$$\rho_k\left(G\right) = \max_x \sum_{i \in V} x_i \qquad (10)$$

при обмеженнях:

$$\sum_{j \in \Gamma(i)} x_j \leq (k-1)x_i + \overline{d}_i(1 - x_i), \quad \forall i \in V, \qquad (11)$$

$$x_i \in \{0, 1\}, \quad \forall i \in V, \qquad (12)$$

де $\overline{d}_i = |\,\overline{\Gamma}(i)\,|$. Спочатку додаватимуться функціонально надлишкові обмеження, одержані домноженням кожного з лінійних обмежень у (11) на ті змінні x_l, які не входять у це обмеження, тобто додаються функціонально надлишкові обмеження першого типу із сімейства $n(n-1)$ обмежень вигляду

$$\sum_{j \in \Gamma(i)} x_l x_j \leq (k-1)x_l x_i + \overline{d}_i(1 - x_i)x_l, \quad \forall i, l \in V, \quad i \neq l. \qquad (13)$$

Далі додаються функціонально надлишкові обмеження другого типу, одержані з лінійних обмежень (11) домноженням на $1 - x_l$, $l = 1, 2, \ldots$. Тут уже можна використовувати $i = l$, бо вони дають нові квадратичні обмеження у формі нерівностей. У результаті одержуємо n^2 обмежень

$$\sum_{j \in \overline{\Gamma}(i)} x_l x_j - \sum_{j \in \overline{\Gamma}(i)} x_j \leq (k-1)x_l x_i + \overline{d}_i(1 - x_i)x_l, \quad \forall i, l \in V. \qquad (14)$$

Для графів G_1 та G_2 з рис. 1 нижче буде показано, що за допомогою функціонально надлишкових обмежень із сімейств (13) і (14) лагранжеві двоїсті оцінки можна зробити точними верхніми оцінками для $\rho_2(G_1)$ та $\rho_2(G_2)$.

Уточнення лагранжевих двоїстих оцінок для максимального 2-плекса

Розглянемо задачу знаходження максимального 2-плекса для графа G_1 з рис. 1, *а*. Вона формулюється у вигляді такої квадратичної задачі:

$$\rho_2\left(G_1\right) = \max_{x \in R^5}(x_1 + x_2 + x_3 + x_4 + x_5) \qquad (15)$$

при обмеженнях:

$$\begin{cases} x_1x_2 + x_1x_5 \leq x_1, \\ x_1x_2 + x_2x_3 \leq x_2, \\ x_2x_3 + x_3x_4 \leq x_3, \\ x_3x_4 + x_4x_5 \leq x_4, \\ x_4x_5 + x_1x_5 \leq x_5. \end{cases} \qquad (16)$$

$$x_i^2 - x_i = 0, \quad i = 1, 2, 3, 4, 5. \qquad (17)$$

Тут $\rho_2\left(G_1\right) = 3$. Водночас лагранжева двоїста оцінка для задачі (15)–(17) дорівнює 3.618034. Вона є менш точною, ніж лінійна верхня оцінка, що виходить у результаті релаксації задачі булевого програмування (10)–(12) і дорівнює 10/3 = 3.33333. Задачі лінійного програмування відповідають такі лінійні обмеження:

$$\begin{cases} x_1 + x_2 + x_5 \leq 2, \\ x_1 + x_2 + x_3 \leq 2, \\ x_2 + x_3 + x_4 \leq 2, \\ x_3 + x_4 + x_5 \leq 2, \\ x_1 + x_4 + x_5 \leq 2, \end{cases}$$

які випливають з обмежень (11).

Як лагранжева двоїста оцінка, так і лінійна оцінка є неточними верхніми оцінками. Покращити лагранжеву двоїсту оцінку, що відповідає задачі (15)–(17), можна, додаючи до квадратичної задачі (15)–(17) функціонально надлишкові обмеження із сімейства (13). Вони побудовані домноженням лінійних обмежень (24) на змінні x_i, $i = 1, \ldots, n$. Послідовне додавання таких обмежень робить лагранжеву двоїсту оцінку більш точною. При послідовному додаванні функціонально надлишкових квадратичних обмежень із сімейства (13), вказаних у таблиці, спостерігається монотонне зменшення оцінки ψ^*. У результаті додавання чотирьох додаткових обмежень верхня оцінка ψ^* стає точнішою за лінійну. А в результаті додавання всіх п'яти обмежень оцінка ψ^* стає точною верхньою оцінкою для $\rho_2\left(G_1\right)$.

Таблиця. **Покращення верхньої оцінки для задачі (15)–(17)**

$+n$	Лінійні обмеження	$\times x_i$	Квадратичні обмеження	ψ^*
+1	$x_1 + x_2 + x_5 \le 2$	$\times x_2$	$x_1x_2 + x_2x_5 \le x_2$	3.56878
+2	$x_1 + x_2 + x_3 \le 2$	$\times x_3$	$x_1x_3 + x_2x_3 \le x_3$	3.52727
+3	$x_2 + x_3 + x_4 \le 2$	$\times x_4$	$x_2x_4 + x_3x_4 \le x_4$	3.42929
+4	$x_3 + x_4 + x_5 \le 2$	$\times x_5$	$x_3x_5 + x_4x_5 \le x_5$	3.21298
+5	$x_1 + x_4 + x_5 \le 2$	$\times x_1$	$x_1x_4 + x_1x_5 \le x_1$	3.00000

Схожа, але дещо інша, ситуація має місце і для задачі знаходження максимального 2-плекса для графа G_2 з рис. 1, *б*. Вона формулюється у вигляді квадратичної задачі

$$\rho_2\left(G_2\right) = \max_{x \in R^5}(x_1 + x_2 + x_3 + x_4 + x_5) \qquad (18)$$

при обмеженнях:

$$\begin{cases} x_1x_2 + x_1x_3 + x_1x_5 \le x_1, \\ x_1x_2 + x_2x_3 \le x_2, \\ x_1x_2 + x_2x_3 + x_3x_4 \le x_3, \\ x_3x_4 + x_4x_5 \le x_4, \\ x_4x_5 + x_1x_5 \le x_5. \end{cases} \qquad (19)$$

$$x_i^2 - x_i = 0, \quad i = 1, 2, 3, 4, 5, \qquad (20)$$

і їй відповідає $\rho_2\left(G_2\right) = 3$. Лагранжева двоїста оцінка для задачі (18)–(20) дорівнює 3.37332 і є точнішою, ніж лінійна верхня оцінка, що дорівнює 3.4. Задачі лінійного програмування відповідають такі лінійні обмеження виду (11):

$$\begin{cases} 2x_1 + x_2 + x_3 + x_5 \le 3, \\ x_1 + x_2 + x_3 \le 2, \\ x_1 + x_2 + 2x_3 + x_4 \le 3, \\ x_3 + x_4 + x_5 \le 2, \\ x_1 + x_4 + x_5 \le 2. \end{cases} \qquad (21)$$

Враховуючи, що лагранжева двоїста оцінка є точнішою за лінійну, може скластися враження, що точну оцінку ψ^* простіше одержати за допомогою додавання функціонально надлишкових обмежень (13), тобто тих, котрі побудовані з лінійних обмежень (21) домноженням на змінні x_i, $i = 1,..., 5$. Однак це не так, і максимальна за точністю лагранжева двоїста оцінка, яку можна досягти на цьому шляху, дорівнює 3.02316. Лише при додаванні функціонально надлишкових обмежень у вигляді (14) можна домогтися того, щоб лагранжева двоїста оцінка стала точною. Наприклад, після додавання до квадратичної задачі (18)–(20) функціонально надлишкового обмеження виду (13)

$$x_1x_1 + x_1x_4 + x_1x_5 - 2x_1 \le 0,$$

яке одержано із п'ятого обмеження (21) домноженням на x_1, оцінка ψ^* стає рівною 3.21576. Після додавання обмежень виду (14)

$$-x_1x_3 - x_1x_4 - x_1x_5 + 2x_1 + x_3 + x_4 + x_5 \le 2,$$

яке отримано домноженням четвертого обмеження з (21) на $(1 - x_1)$, оцінка ψ^* стає рівною 3.01571. А після додавання ще й обмежень виду (14)

$$-2x_1x_2 - x_2^2 - x_2x_3 - x_2x_5 + 2x_1 + 4x_2 + x_3 + x_5 \le 3,$$

яке одержано домноженням першого обмеження з (21) на $(1 - x_2)$, оцінка ψ^* стає точною й дорівнює 3.0.

Висновки

Із прикладів квадратичних задач (15)–(17) та (18)–(20) для знаходження максимальних 2-плексів графів G_1 та G_2 видно, що функціонально надлишкові обмеження виду (13) та (14) відіграють значну роль для поліпшення точності лагранжевих двоїстих оцінок. Так, з їх допомогою лагранжеву двоїсту оцінку ψ^* можна зробити точною для $\rho_2\left(G_1\right)$ та $\rho_2\left(G_2\right)$. Однак це не означає, що для будь-якої квадратичної задачі (1)–(3) за допомогою обмежень виду (13) та (14) можна домогтися того, щоб лагранжева двоїста оцінка була точною. Для підвищення точності лагранжевих двоїстих оцінок можна використовувати й інші способи побудови функціонально надлишкових обмежень [7].

Робота виконана за підтримки НАН України, проект 0117U000327.

Список літератури

1. Стецюк П. І. Квадратична задача для максимального *k*-плекса в неорієнтованому графі / П. І. Стецюк, Т. О. Бардадим, В. І. Ляшко // Журнал обчислювальної та прикладної математики. – 2017. – № 1 (124). – С. 71–78.
2. Balansundaram B. Clique Relaxations in Social Network Analysis: The Maximum *k*-plex Problem / B. Balansundaram, S. Butenko, I. V. Hicks // Operations Research. – 2011. – Vol. 59, no. 1. – P. 133–142.
3. Guo J. A more relaxed model for graph-based data clustering: *s*-plex cluster editing / J. Guo, C. Komusiewicz, R. Niedermeier, J. Uhlmann // SIAM J. Discrete Math. – 2010. – Vol. 24, no. 4. – P. 1662–1683.
4. McClosky B. Combinatorial algorithms for the maximum *k*-plex problem / B. McClosky, I. V. Hicks // J. of Combinatorial Optimization. – 2012. – Vol. 23, no. 1. – P. 29–49.
5. Pattillo J. On the maximum quasi-clique problem / J. Pattillo, A. Veremyev, S. Butenko, V. Boginski // Discrete Applied Mathematics. – 2013. – Vol. 161, no. 1–2. – P. 244–257.
6. Seidman S. B. A graph theoretic generalization of the clique concept / S. B. Seidman, B. L. Foster // J. of Math. Sociology. – 1978. – Vol. 6. – P. 139–154.
7. Shor N. Z. Nondifferentiable Optimization and Polynomial Problems / N. Z. Shor. – London/Boston/Dordrecht : Kluwer Academic Publishers, 1998. – 394 p.
8. Shor N. Z. Dual Solution of Quadratic-Type Problems by *r*-algorithm (subroutine DSQTPr) / N. Z. Shor, P. I. Stetsyuk // Abstracts of the Second International Workshop "Recent Advances in Non-Differentiable Optimization" (October, 1–4, 2001, Kyiv, Ukraine). – Kyiv, 2001. – P. 36.

P. Stetsyuk, V. Lyashko, T. Bardadym

PROPERTIES OF THE QUADRATIC PROBLEM ABOUT MAXIMUM *k*-PLEX IN UNDIRECTED GRAPH

The paper considers properties of upper bounds in the quadratic problem about maximum k-plex in undirected graph. The notion of k-plex is introduced to make the notion of clique to be less restrictive. These notions are widely used in sociology to identify specific subgroups of the population. They also can be used to clusterize data, to optimize information flows in networks, to find targeted subgroups for advertising and promotion, etc.

In a general case, optimization problems of finding maximum cliques and k-plexes are NP-hard. For this reason, in the maximum clique problem, finding dual bounds is a widely used approach. These bounds may be used directly, especially if the exact bound can be found, or they may be an integral part of other algorithms similar, for example, to the branch-and-bound algorithm.

The relationship between the quadratic problem for 1-plex and a known formulation of the quadratic problem for finding the maximum clique in a graph is analyzed. Lagrangian dual quadratic bounds for the simplest quadratic problems are reported; it is shown that these bounds can be improved by adding functionally superfluous constraints. This approach was proposed by N. Z. Shor for the maximum clique problem and also used in polynomial optimization.

Different ways to generate functionally superfluous constraints are proposed. On the example of maximum 2-plex problem different formulation of the resulting quadratic programming problem are discussed and compared. This does not mean, however, that the considered superfluous constraints give an exact dual bound in any quadratic problem. To improve Lagrangian dual quadratic bounds, other ways of generation of superfluous constraints can also be used.

Keywords: maximum *k*-plex, maximum clique, quadratic problem, Lagrangian dual bounds, functionally superfluous constraints.

Матеріал надійшов 18.05.2017